dprl@10: The Document and Pattern Recognition Lab's First 10 Years

Richard Zanibbi

Biography

Dr. Zanibbi joined the Department of Computer Science at RIT in 2006, where he directs the Document and Pattern Recognition Lab (DPRL – www.cs.rit.edu/~dprl). His research interests include Document Recognition, Information Retrieval, and Human-Computer Interaction. Funding for his research has been provided by Xerox, Google, the NSF, and the Alfred P. Sloan foundation. The DPRL has an international profile, and the lab has hosted research visitors from around the world. Dr. Zanibbi has held Visiting Professor appointments at the University of Nantes (France) and University of Waterloo (Canada).

Dr. Zanibbi received his PhD in Computer Science from Queen's University (Canada), and was an NSERC Postdoctoral Fellow at the Centre for Pattern Recognition and Machine Intelligence in Montreal before coming to RIT. He has published in leading international journals and conference proceedings (e.g. TPAMI, CVPR, IJDAR, and SIGIR), and is best known for his work on the recognition and retrieval of mathematical notation, Video CAPTCHAs, and detecting text in images of natural scenes.

He co-chaired the 2012 and 2013 SPIE Document Recognition and Retrieval (DRR) conferences, and is co-chairing the upcoming International Conference on Frontiers in Handwriting Recognition (ICFHR 2018) being held in Niagara Falls. Dr. Zanibbi is the Communications Officer for the International Association on Pattern Recognition (IAPR) TC-11 (Reading Systems), and serves on a variety of international conference program committees and journal editorial boards (e.g., ICDAR, ICFHR, IJDAR, and IET Computer Vision).

The Goal: Improve algorithms and tools for recognition and retrieval of information in documents, images, audio and video.

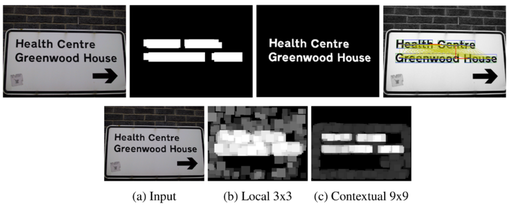
The Document and Pattern Recognition Lab (DPRL) started in the summer of 2007, and this year marks the 10th anniversary of the lab. This poster summarizes the accomplishments of DPRL students during the lab's first decade. The DPRL is collaborative and student-centered, with its primary goal being to make contributions to the international research community. This goal is used to structure how students are trained: students in the lab learn about representing problems and their solutions mathematically, research programming, compiling literature reviews, experimental design, collecting and analyzing results, collaboration, academic writing, and giving presentations. Our work has been disseminated widely - in addition to presentations that I have given, in the last three years DPRL students have given presentations at leading conferences including NTCIR 2014 in Japan, DRR 2015 in San Francisco, SIGIR 2014 in the Gold Coast, Australia, SIGIR 2016 in Pisa, Italy, SIGIR 2017 in Tokyo, Japan, and CVPR 2016 in Las Vegas.

Video CAPTCHA
(Kurt Kluever, CUPS 2009)

- Identify computers ('bots') vs. humans. One 'correct' word to pass. Words stemmed (via Porter), synonyms allowed.
- 'Correct' answers generated automatically from tags on 'similar' YouTube videos (by cosine similarity). Filter tags with frequency > t %.
- In an experiment, students passed 90% of challenges, vs. 13% break rate using most common tags.
- **Innovations:** First video CAPTCHA. Random walk-based video sampling, tag set expansion and filtering methods.

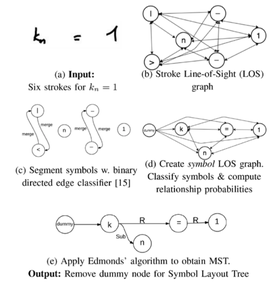
Text-Conv: Text Detection in Natural Scenes
(Siyu Zhu, CVPR 2016; David Syner, Bo Ding, Kardo Aziz)

- Cascaded AdaBoost-based detector that obtained state-of-the-art results for ICDAR 2015 benchmark without OCR or conv. nets.
- **Innovations**: patch context (8-neighborhood); sampling techniques (coarse to fine; graph constraints); MST word segmentation; extended convolutional k-means [Coates & Ng] for patch learning.



(a) Input (b) Local 3x3 (c) Contextual 9x9

Handwritten Math Recognition
(Lei Hu, ICFHR 2016; Ouyang, Davila, Condon, Ravi)



(a) **Input:** Six strokes for $k_n = 1$

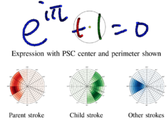
(b) Stroke Line-of-Sight (LOS) graph

(c) Segment symbols w. binary directed edge classifier [15]

(d) Create *symbol* LOS graph. Classify symbols & compute relationship probabilities

(e) Apply Edmonds' algorithm to obtain MST. **Output:** Remove dummy node for Symbol Layout Tree

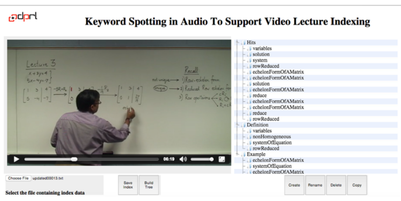
- Efficient parser for math written on a tablet.
- **Innovations:** LOS constraints; Parzen shape context features; accurate segmentation & structure without OCR



Expression with PSC center and perimeter shown

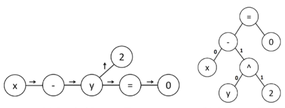
Parent stroke Child stroke Other strokes

Keyword Spotting in Lecture Audio
(Miller, Kanadje, Agarwal, et al., PR Letters, 2016; Miller-Jacobson)



- Instructor searches and indexes lecture videos with queries recorded on a laptop. Interface to categorize and organize contents.
- Unsupervised MFCC + Segmental DTW [Park & Glass]. Average 70% of top 10 hits were correct; within lecture queries 80%; 90% for lapel mic-recorded MIT lectures.
- **Innovations:** whitening of MFCC vectors, reducing strength of low frequencies on laptop recordings, noise in audio recording.

Tangent Formula Search Engine
(Davila, SIGIR 2016 & 2017; Schellenberg, Stalnaker, Pattaniyil)



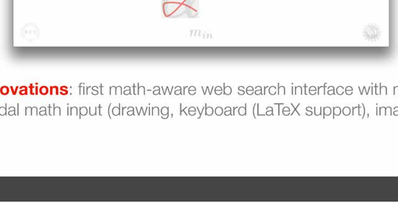
(a) Symbol Layout Tree (b) Operator Tree

Figure 1: Tree representations for $x - y^2 = 0$

	Core	Matching
1.	$O(mn \log m)$	$O(mn \log m)$
2.	$O(mn)$	$O(VE \log V)$
3.	$O(mnp)$	$O(VE \log V \log(VC))$
4.	$O(m + \log n)$	$O(Tm) = O(n^2 m \log n)$
5.	$O(m\sqrt{n} \log n)$	$O(nk \log k)$

- Symbol pair path matching, re-ranking by best query match.
- **Innovations:** symbol pair-based model. Unification and wildcard support in rerank. Near state-of-art results for NTCIR 12 in Tokyo; orders of magnitude faster than 'best' system (real-time)

m_{in} Math Search Interface
(Sasarak, Hart, Pospesel, Stalnaker, Hu, LiVolsi, Zhu, HCIR 2012; Wangari, SIGIR 2014; Orakwue, CICM 2015)



- **Innovations**: first math-aware web search interface with multi-modal math input (drawing, keyboard (LaTeX support), images)